MITRE ENGENUITY | Center for Threat Informed Defense

MTRENCT0053
MITRE ENGENUITY TECHNICAL REPORT

Cloud Analytic Development Blueprint

Best Practices and Lessons Learned for Developing Cloud Analytics

Author(s):
Michael Butt
Ingrid Skoog
Desiree A. Beck, PhD

July 2022

MITRE ENGENUITY | Center for Threat Informed Defense

Abstract

The Center for Threat-Informed Defense's Cloud Analytics project researched and developed best practices and guidance to help defenders improve their ability to detect adversary behaviors in today's complex cloud environments. This cloud analytic development blueprint document characterizes our research, the challenges we faced, and the resources that we developed.

Creating effective cloud analytics for different environments proved challenging. The team encourages sharing cloud analytics within the community to enable defenders to learn from each other and build on the lessons learned and experience of others.

The accompanying GitHub repository contains the Sigma rules created during the project, as well as supporting project resources. Future opportunities include creating a Sigma converter to convert Sigma rules to Google Big Query, enhancing Elastic Common Schema to expand mapping scope, and expanding support for Sigma correlation rules within pySigma.

TABLE OF CONTENTS

List of Figures

List of Tables

1 Introduction

There is a global migration of information technology data, services, and software from on-premises, legacy environments to the cloud. Today, enterprises rely on a complex collection of resources hosted across a variety of cloud environments. Defenders do not have the same level of experience performing detection in cloud environments and often lack the same level of visibility when compared to legacy environments. The Center for Threat-Informed Defense launched the Cloud Analytics project to research and develop best practices to help defenders improve their ability to detect adversary behaviors in today's complex cloud environments.

Cyber analytics are a data-centric approach to cybersecurity based on analyzing large volumes of event data collected from heterogenous sources. High quality analytics can quickly identify instances of malicious behavior in a sea of data, reducing the time to detection and focusing the defenders' attention on the most important activities.

The analytic developers must carefully balance the impact of false positives since analytics with low signal-to-noise ratios require more staff time to process and draw attention away from true threats. Projects such as Sigma[1] provide a structured format for detection analytics that can be translated to platform-specific queries for platforms such as Splunk, CrowdStrike, Google Chronicle, and Microsoft Sentinel, among others.[2]

The project focused on Microsoft Azure and Google Cloud Platform (GCP), but the results generalize to other cloud platforms. The project leveraged the MITRE ATT&CK®[3] for Cloud Matrix**Error! Bookmark not defined.** as a reference for cloud-specific a dversary behavior. The deliverables include a set of cloud analytics for key tactics, techniques, and procedures (TTPs) as well as best practices and lessons learned during the project. This blueprint document provides guidance on the processes, challenges, and workflows that can be used by defenders to create their own cloud analytics. The intended audience is creators of cyber analytics for cloud environments, but analytics users can also benefit from this document.

Figure 1. Learn-by-doing project approach

This document is grounded in the practical experience of iteratively developing, testing, and refining cyber analytics in real cloud environments.

2 Analytic Development

2.1 Approach

Creating a cyber analytic is a complex task that requires balancing false positive risk and detection accuracy of the intended target behavior. The Cloud Analytics project experimented with two approaches, the initial adversary-focused approach that encountered challenges, and a second, targeted, technique-focused approach that built on the lessons learned from the initial approach. We would recommend the technique-focused approach for future efforts, as technique-focused approach provides a more focused workflow focused on the target technique throughout the emulation, detection, and evaluation processes.

2.1.1 Adversary-Focused Approach

The initial project approach consisted of identifying cloud adversaries, developing emulation plans, executing those emulation plans, and sifting through the results to find meaningful signals to build analytics on.

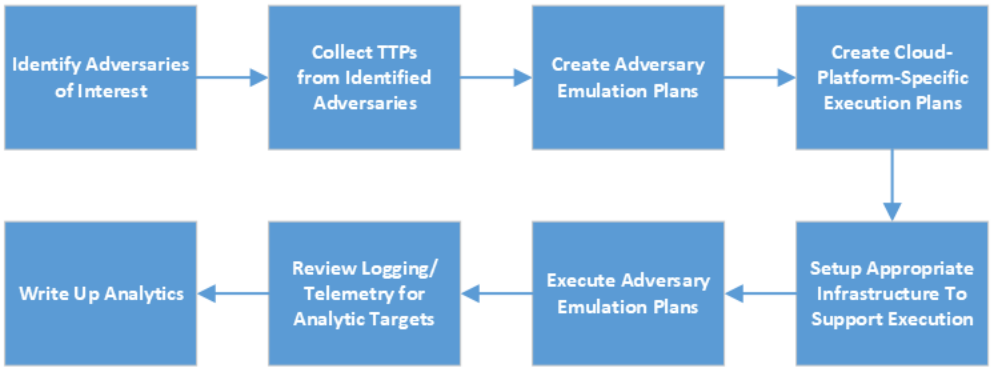


Figure 2. Initial approach to developing cyber analytics

This approach proved challenging and produced disappointing results. The first issue was the limited amount of open-source intelligence for cloud-specific adversaries. The Cloud Analytics project team initially focused on TeamTNT, a threat actor known for cryptocurrency mining campaigns on cloud infrastructure. To produce a well-rounded adversary profile, the project team augmented the TeamTNT adversary emulation plan with additional ATT&CK TTPs for exploiting container-based platforms. The complexity of the environment combined with the breadth of TTPs generated an overwhelming amount of log events. It is difficult to separate the evidence of reconnaissance from the background noise of other normal benign events. Cloud services continuously update configuration maps, leases, leader elections for clusters, and more. Many cloud services perform API requests on service metadata. Distinguishing benign API queries

from malicious events is challenging. Under ideal conditions with minimal cloud assets, cloud platforms such as Azure and GCP can produce hundreds or thousands of log events per hour. As a result, distinguishing useful signals from noise was a bit like searching for a needle in a haystack.

2.1.2 Technique-Focused Approach

In contrast, the technique-focused approach started with the identification of a small set of cloud native TTPs. Compared with the TeamTNT emulation plan, this approach reduced complexity and logging noise. As a bonus, it also simplified the mapping of analytics to ATT&CK techniques, since each analytic was paired with a specific TTP at the start of the process.

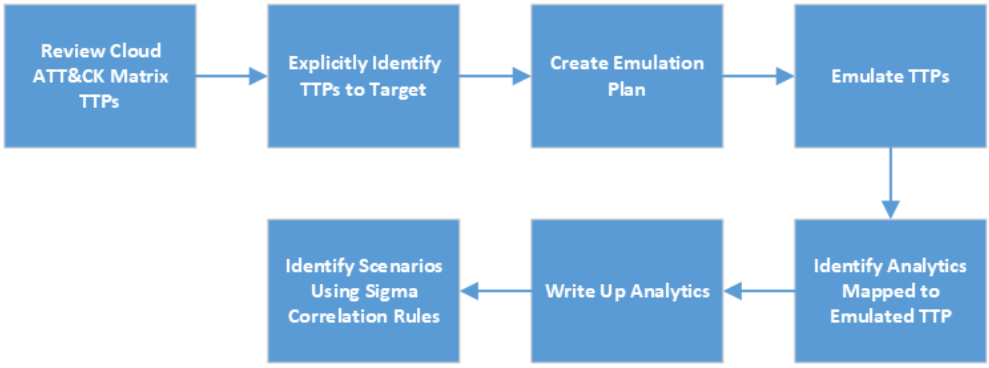


Figure 3. Targeted approach to developing cyber analytics

2.2 Analytics

Analytics for the project take the form of Sigma rules. Sigma is a common language for defenders to define log events of interest. Conversion engines, such as sigmac or PySigma, translate rules from the Sigma language into a platform specific query, such as Google Chronicle, Microsoft Sentinel, Splunk, or Elastic Search. This standard is powerful because it enabled the Cloud Analytics team to create one set of vendor-agnostic analytics that can be deployed in a variety of environments.

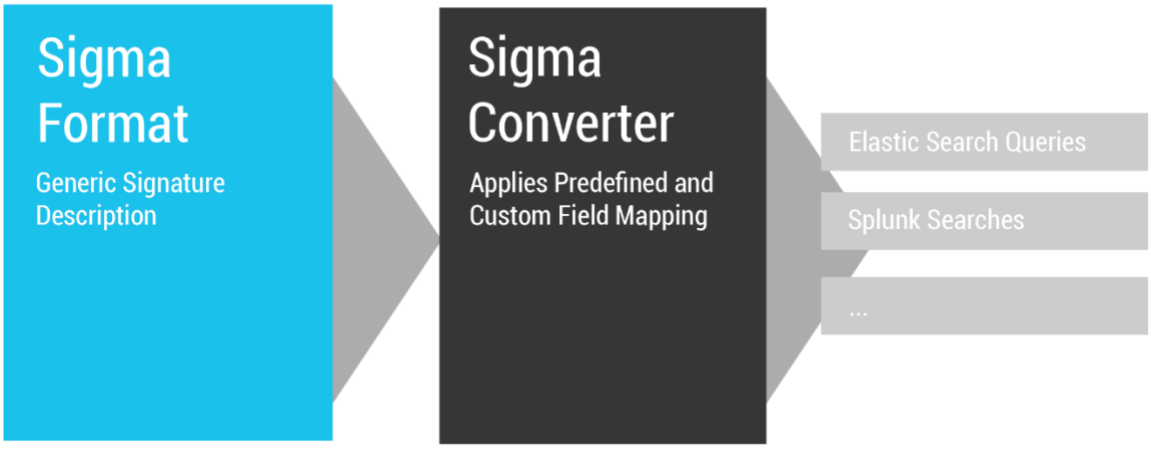


Figure 4. Sigma rule conversion workflow[4]

The project team created detection rules designed to detect specific events mapping to specific ATT&CK techniques. These individual detection rules are prone to false positives and should be considered building blocks of a layered detection program. In 2022, the Sigma project released a new feature for creating correlation rules—rules that match multiple, related events. Due to timing, a small set of preliminary correlation rules were integrated into the Cloud Analytics project. This is an area of future research, and the applications of correlation rules are discussed in the next section.

2.2.1 Correlation Analytics

Sigma Correlations[5] are a new addition to the Sigma standard that builds on individual detection rules for "the expression of aggregations and relationships between rules." The correlation concept allows for expressing temporal relationships between detection rules and grouping detections by attributes such as hostname or username.

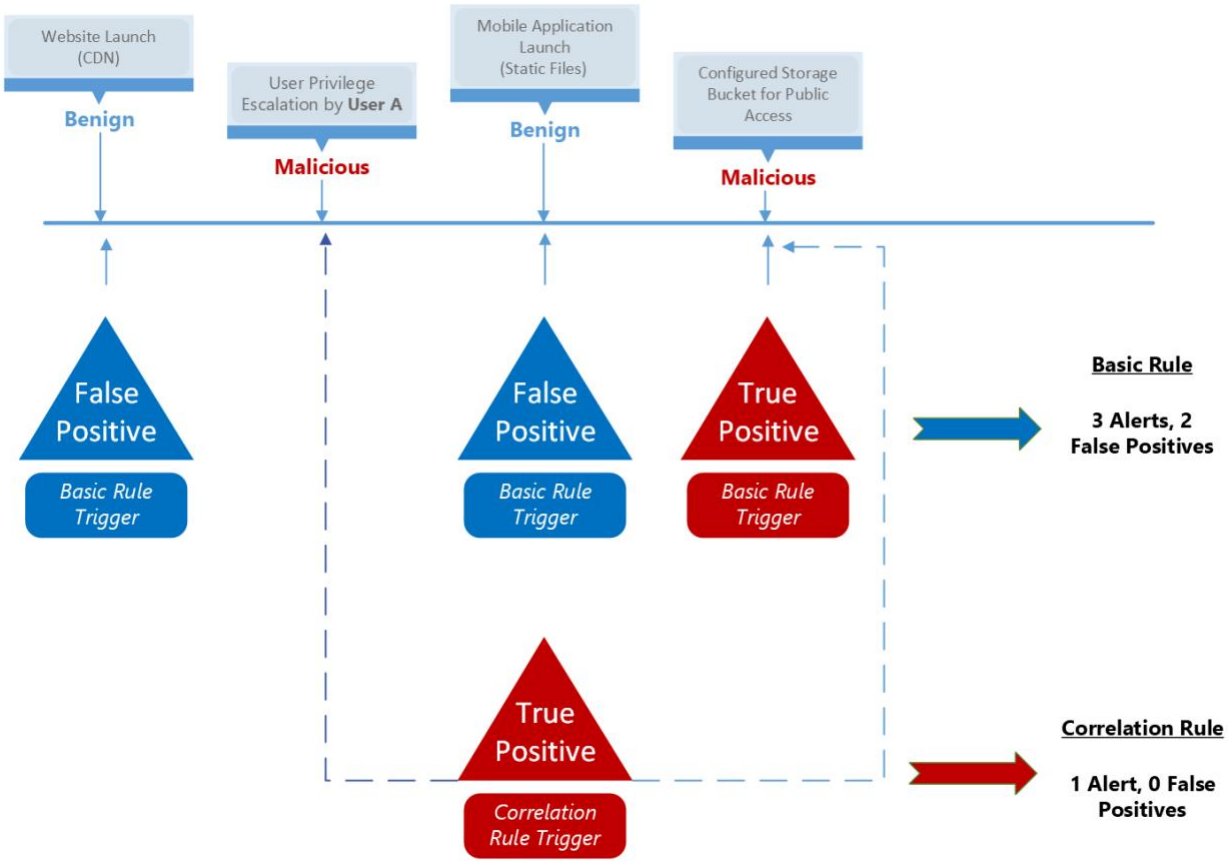


Figure 5. Diagram comparing standard detection rules vs. correlation rules

For example, Figure 5 shows a basic analytic that detects when a cloud storage container is modified to allow public access. This analytic may generate false positives if, for example, a new website is launched, and its storage bucket is switched from private to public as part of the release. The correlation rule depicted above detects the specific sequence of 1) a guest user being granted administrator access, and then 2) that same guest user modifying a storage container to allow public access. The correlation rule provides a stronger signal of potential malicious activity.

2.2.2 Cross-Cloud Analytics

Cloud platforms (such as Microsoft Azure and GCP) have significant differences in terms of organizational structure and resource organization. However, there are high-level cloud concepts that can usually be mapped to the specifics of each platform. Theoretically, Sigma is a common language that maps to these cross-cloud concepts and can be converted into platform-specific rules. In practice, there are significant limits at play when creating cloud-agnostic Sigma rules. The concepts are generalizable (creation of additional resources, autoscaling usage, modification of access controls, and cross-cloud service references are immensely helpful in identifying equivalent

services[6]), but service names, logging architectures, and detection queries all vary significantly between platforms. These variations require the creation of distinct rules for each cloud platform. Cross-cloud service references are immensely helpful for mapping out equivalent services across cloud platforms.[7]

Other projects, such as Elastic Common Schema (ECS), are working to provide a common schema to map event data across platforms. For example, an ECS object will store information about a user under the "User" field set regardless of the source. Azure may provide the user metadata via the User Principal Name, and GCP may provide the user email under a principalEmail field; however, the ECS processor for both log sources would map the data to the appropriate ECS attributes. The common schema allows for analysis across all data sets.

3 Case Study

To further examine the analytic development process, we will work through an example using the targeted approach discussed earlier.

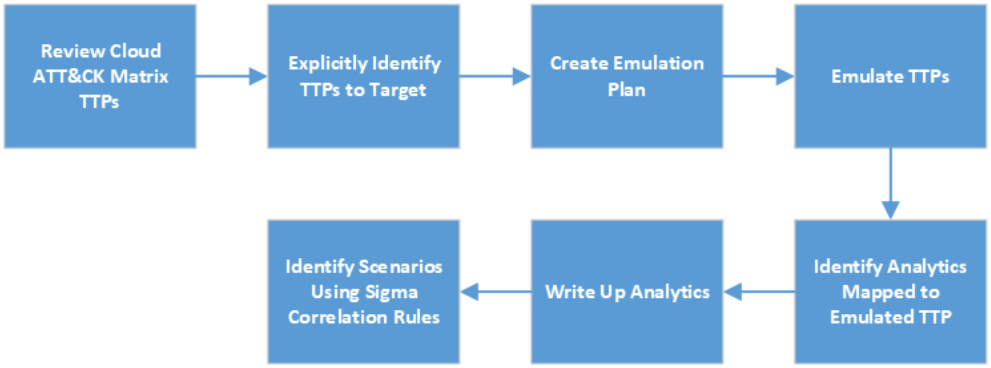


Figure 6. Targeted approach to cyber analytic development

3.1 Identifying a TTP

We start with reviewing the appropriate ATT&CK Matrix for your area of interest. As previously mentioned, this project used ATT&CK Cloud Matrix.[8]

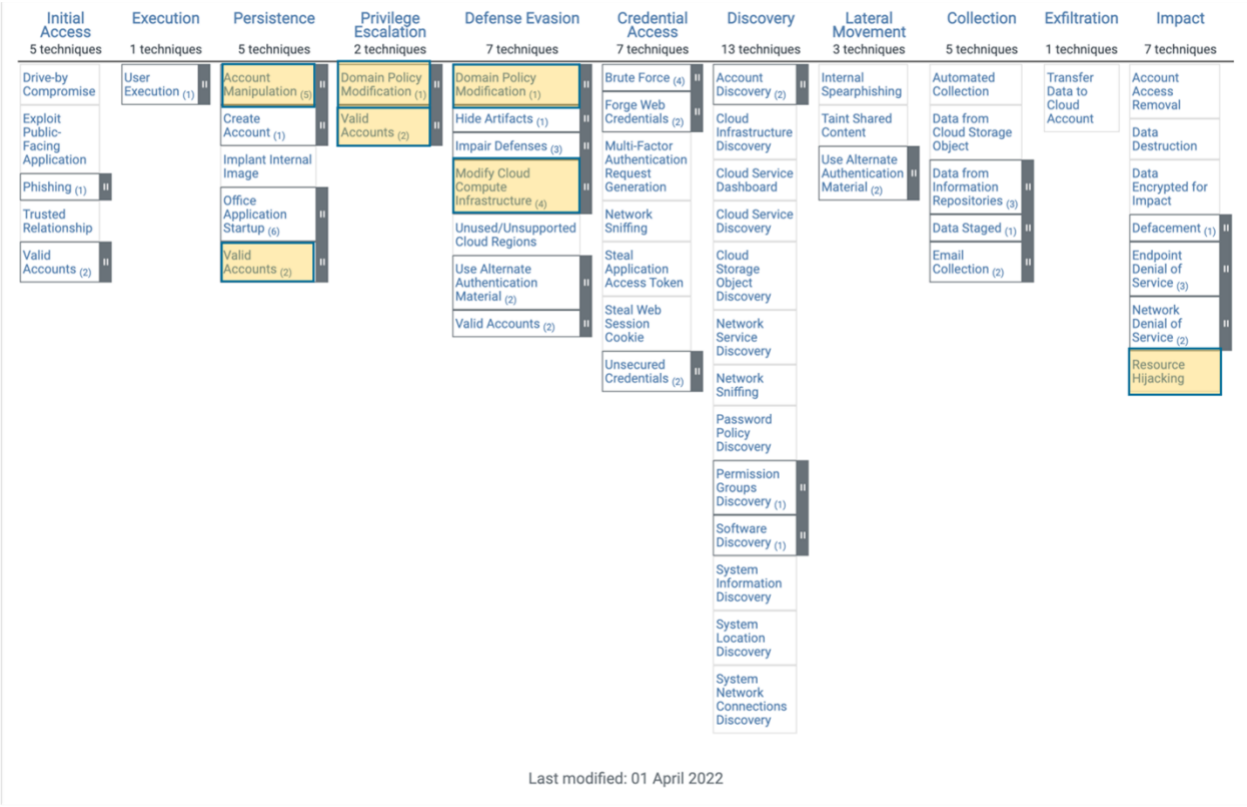


Figure 7. ATT&CK Cloud Matrix

For this example, we select T1578: Modify Cloud Compute Infrastructure[9]. While the T1578 TTP covers a wide range of activities, we will focus on modification of storage buckets to allow public access. The full ATT&CK entry for T1578 can be found in Figure 8.

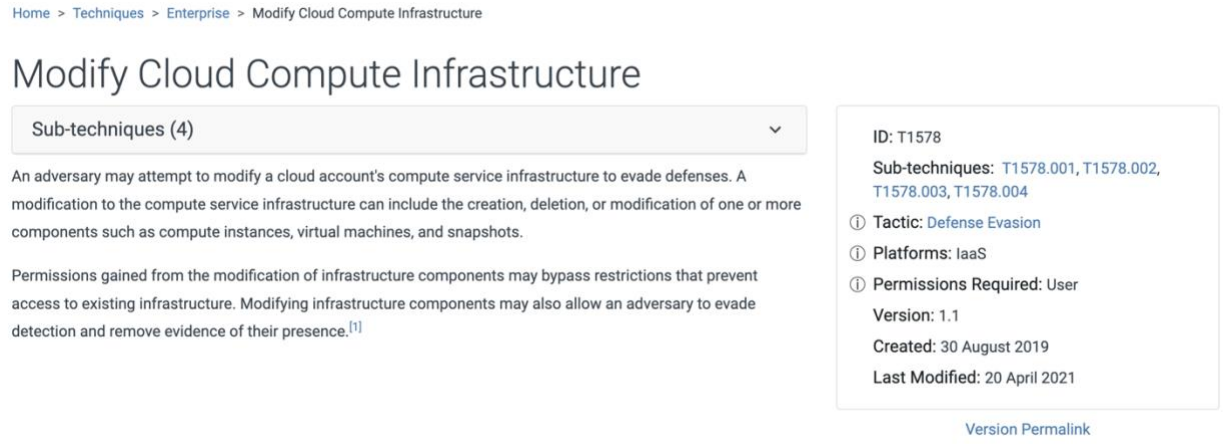


Figure 8. ATT&CK TTP T1578

3.2 Adversary Emulation

Now that we have identified a specific TTP and resource target, the next step is adversary emulation. Adversary emulation uses TTPs based on real world activity, leveraging ATT&CK, to provide a controlled execution of specific TTPs. In this case, the following commands will emulate modifying a cloud storage container to allow for public, unrestricted access. Both Azure and GCP provide documentation on using their command line interface (CLI) tools to make a storage container public.

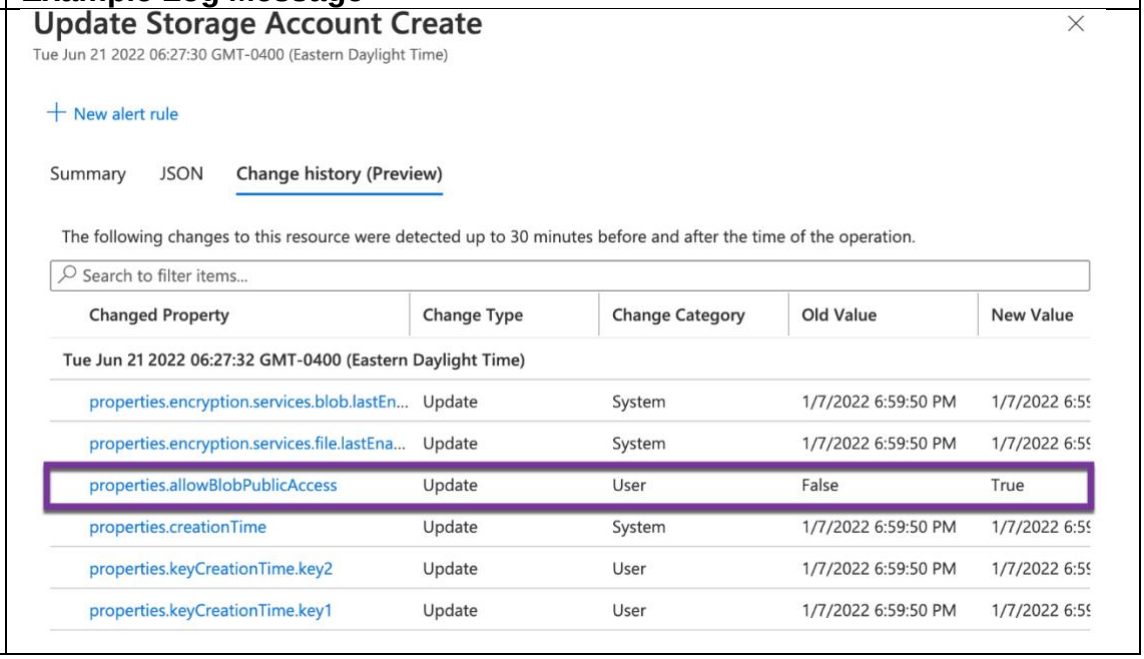
The respective commands for Azure and GCP are below.

Service	Command
Azure[10]	az storage account update \ --name <storage-account> \ --resource-group <resource-group> \ --allow-blob-public-access true
GCP[11]	gsutil iam ch allUsers:objectViewer gs://*BUCKET_NAME*

Figure 9. Commands to make a storage bucket public on Azure and on GCP

3.3 Identify Analytic and Map to TTP

Once the emulated attack has executed, review the appropriate documentation and logs to identify any significant events that resulted from it. The following table provides examples of a storage blob being set to public for Azure and GCP.

Service	Example Log Message
Azure	

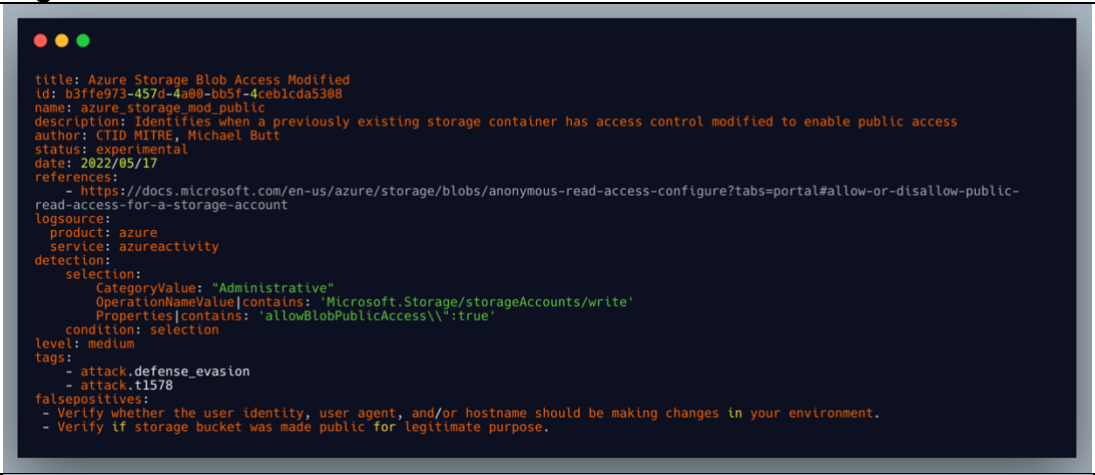
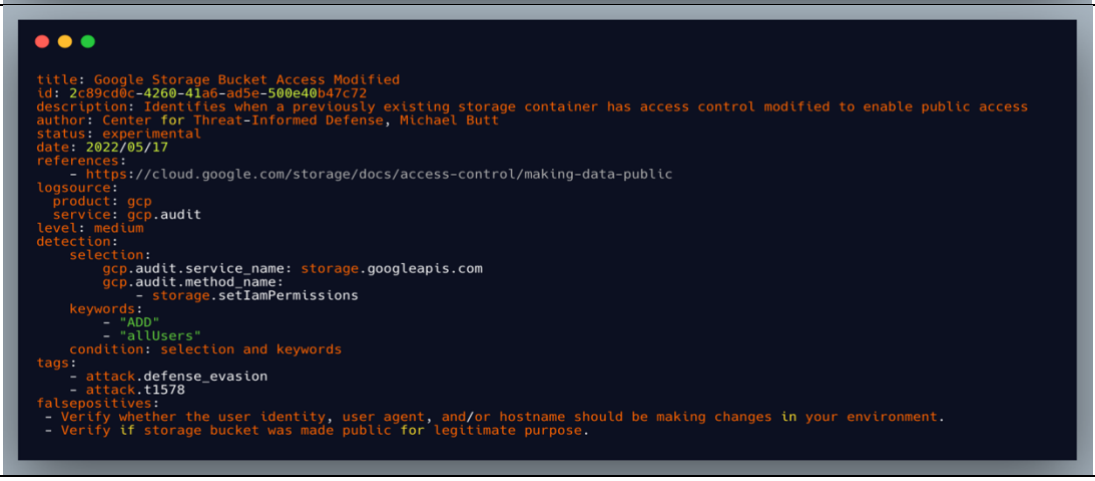
Service	Example Log Message
GCP	

Figure 10. Examples of storage blob becoming public for Azure and GCP

For the mapping component, as the targeted approach starts with an ATT&CK TTP, the ATT&CK mapping for the TTP is already complete. In this case, the analytic will be mapped to T1578.

3.4 Write-up Analytic

The write-up component involves compiling the information gathered in previous steps into a Sigma rule. The Sigma project provides useful documentation, as well as third-party blogs and books.[12] Sigma is YAML Ain't Markup Language (YAML) based and provides a schema. For creating/editing Sigma rules, the Visual Studio Code extension sigma[13] is useful in autocompleting tags, providing code snippets, auto generating universally unique identifiers (UUIDs), and performing preliminary static analysis.

Service	Sigma Rule
Azure	
GCP	

3.5 Validate Analytic

Throughout this section, we worked through a case study of identifying a TTP, emulating the TTP within a cloud environment, and creating a Sigma analytic to identify the specific TTP. The next step in the process is to convert the Sigma rule into a platform-specific query for your supported environment, execute the adversary behavior again, and validate the detection in your environment.

Your goal during this process is to continue to refine the analytic, as analytics should go through a rigorous evaluation process prior to being considered stable. For example, all analytics submitted by the community to the open-source Sigma project are initially considered experimental and take roughly 12 months of community usage to be considered stable. Analytics in your environment should be re-evaluated regularly to ensure they meet your organization's requirements for detection sensitivity and specificity.

4 Infrastructure Setup

A properly configured cloud infrastructure environment is the first requirement for developing cloud analytics. If an environment is not readily available, setting up test infrastructure can be a non-trivial step requiring business funding and approvals to provision the required resources. It is important that the provisioned environment is not overly restrictive—in permission access, feature availability, and metric and logging collection—to serve the desired purpose of cyber analytics research and development. The Cloud Analytics project experienced these issues first-hand: the novelty of the project and cybersecurity concerns created bureaucratic hurdles and delays before the project could begin in earnest. The solution involved setting up an isolated set of cloud environments, which is discussed in detail below.

4.1 Infrastructure Requirements

Emulating attacks directly against cloud infrastructure is a new concept for many organizations. To minimize approval overhead, the project team identified the following requirements for a cloud test range.

- Full administrator permissions over the cloud account(s)
- Separate and standalone cloud service account(s), with no connection to production corporate environments
- Sufficient budget for cloud service usage

4.2 Architecture

The required infrastructure components may vary depending on the specific cloud platform used, as well as the goals of the project. For the Cloud Analytics project, the following components were used for the base deployment. Note that some adversary emulation steps may create additional infrastructure. Both Microsoft Azure and GCP services were used during the project for perspective on cross-cloud analytic development.

- Windows Virtual Machine (VM)
- Linux VM
- Container-based application deployment (Azure Container Instances, GCP Compute)
- Secret Manager (Azure Vault, GCP Secret Manager)
- Storage Service (Azure Blob Storage, GCP Storage)
- User Management (Azure Active Directory, GCP Identity and Access Management)
- Virtual Networks resource communication
- Auditing infrastructure for telemetry storage
- Elastic Cloud cluster for log ingestion and analysis

4.3 Deployment

When possible, automated Infrastructure as Code tools, primarily Terraform, were used to define, provision, deploy, and maintain infrastructure. Terraform uses a declarative syntax for specifying cloud resources such as virtual machines and virtual networks. Terraform makes deployment of cloud infrastructure simple and repeatable, immensely reducing the labor required for managing cloud infrastructure.

4.3.1 Cloud Service Rules of Engagement

Cloud services typically have rules of engagement for performing emulated attacks and pen testing against cloud infrastructure. References to policies for different cloud providers are shown in **Table 1**.

Table 1. Cloud Provider Rules of Engagement

Cloud Provider	References
Amazon Web Services (AWS)	14
Google Cloud Platform (GCP)	15, 16
Microsoft Cloud (including Microsoft Azure)	17

An incorrectly provisioned account may result in one or more of the following issues:

- Inability to create/modify user accounts during attack emulation for different scenarios
- Inability to view **all** logged account activity
    - Most corporate cloud configurations limit visibility of logging and metrics to resources under your control.
    - As a result, you will have limited visibility to events logged by protected resources.
    - For example, attempts to compromise secret storage services, such as Azure KeyVault, will not be viewable if the user viewing the logs does not have access to the KeyVault resource.
- Triggering internal security processes that identify malicious activity
    - If the organization perceives the adversary emulation as a real threat, this will likely result in disabling/removal of related accounts and infrastructure.

5 Summary

Creating effective analytics for different cloud environments proved challenging for the team.

The team encourages sharing any created cloud analytics within the community. Sharing allows defenders to learn from each other and build on the lessons learned and

experience of others. As part of the project, the team is sharing the following resources, as well as a set of new cloud analytics targeting the Azure and GCP platforms.

The accompanying GitHub repository contains resources developed as part of this project. The following two tables outline the resources available in the GitHub repository.

Cloud Analytics – Sigma Rules github.com/center-for-threat-informed-defense/cloud-analytics	
Resource	**Summary**
Analytics	Analytics, in Sigma format, generated for the project. Analytics included for both Azure and GCP platforms. Documentation on getting started with creating Sigma rules
Adversary Emulation	Adversary emulation plan in Caldera Adversary Emulation Library format, as well as documentation on running adversary emulation plans through Caldera.
Support Resources	Content created during project, but not explicitly part of the final deliverable. Includes adversary emulation research, terraform modules, and mapping docs related to the Google CSA ATT&CK mapping effort.

Analytic	Description	ATT&CK TTP	Cloud Platform
Autoscaling Threshold Exceeded	Sigma correlation rule that identifies when the number of instances in the resource group is greater than the threshold	Resource Hijacking (T1496)	Azure
Guest User Privilege Escalation	Identifies when a guest user has privileges escalated to Global Administrator.	Valid Accounts (T1078)	Azure
Guest User Privilege Escalation then Storage Blob Access Modified	Sigma correlation rule that identifies the sequence of events when privileges of a guest user are escalated, and the same guest user makes a storage container for public access.	Valid Accounts (T1078), Modify Cloud Compute Infrastructure (T1578)	Azure
Role Elevated Outside of PIM	Identifies when a privileged role assignment has been made outside of the Privileged Identity Management tool.	Domain Policy Modification (T1484)	Azure
Service Principal Privilege Escalation	Identifies when a service principal has privileges escalated to Global Administrator.	Valid Accounts (T1078)	Azure
Storage Blob Access Modified	Identifies when a previously existing storage container has access control modified to enable public access.	Modify Cloud Compute Infrastructure (T1578)	Azure

Multi-Factor Authentication Failure Threshold Exceeded	Identifies when a user has failed multifactor authentication within a time window more than a pre-defined threshold.	Credential Access: Multi-Factor Authentication Request Generation (T1621)	Azure
Autoscaling Threshold Exceeded	Sigma correlation rule that identifies when autoscaling events have exceeded a pre-threshold.	Resource Hijacking (T1496)	GCP
Permissions Granted Over Service Account	Identifies when permissions granted to principal to impersonate or create keys for a service account.	Domain Policy Modification: Domain Trust Modification (T1484.002)	GCP
SSH Key Added	Identifies when an SSH key is added to an instance.	Account Manipulation: Additional Cloud Credentials (T1098.001)	GCP
Google Storage Bucket Access Modified	Identifies when a previously existing storage container has access control modified to enable public access.	Modify Cloud Compute Infrastructure (T1578)	GCP
Google VPC Service Controls Violation for Storage Bucket Access	Identifies when a Storage Bucket access attempt has been blocked by VPC Service Controls.	Valid Accounts: Cloud Accounts (T1078.004), Exfiltration: Transfer Data to Cloud Account (T1537)	GCP
Workspace Login Marked Suspicious (GCP)	Identifies when a workspace login is marked suspicious.	Valid Accounts: Cloud Accounts (T1078.004)	GCP
Workspace User Added to Privileged Group	Identifies when a user is added to a privileged group.	Valid Accounts: Cloud Accounts (T1078.004), Account Manipulation: Additional Cloud Roles (T1098.003)	GCP

Future opportunities related to the project include creating a Sigma converter to convert Sigma rules to Google Big Query. Sigma is in the process of switching to the pySigma toolset for rule conversion as the sigmac tool has been deprecated. There are ongoing discussions with the Sigma maintainers and will continue work to coordinate such efforts. Other potential follow-on research opportunities would include enhancements to Elastic Search ECS common data structure to allow for expanded mapping scope and expanding support for Sigma correlation rules within pySigma.

Appendix A Terminology

Term	Description
Cyber Analytics Repository	The MITRE Cyber Analytics Repository is a knowledge base of analytics developed by MITRE based on the MITRE ATT&CK adversary model.
MITRE ATT&CK	MITRE ATT&CK is a globally accessible knowledge base of adversary tactics and techniques based on real-world observations.
ATT&CK Cloud	ATT&CK Cloud corresponds to tactics and techniques representing the MITRE ATT&CK Matrix for Enterprise covering cloud-based techniques.
Sigma	Sigma is a generic and open signature format that allows you to describe relevant log events in a straightforward manner. The rule format is very flexible, easy to write, and applicable to any type of log file.
ATT&CK Evaluations	The ATT&CK Evaluations program, operated by MITRE Engenuity, provides open and fair evaluations based on ATT&CK.
Infrastructure as a Service	Infrastructure as a Service are online services that provide high-level application programming interfaces (APIs) to dereference various low-level details of underlying network infrastructure, like physical computing resources, location, data partitioning, scaling, security, and backup.
Infrastructure as Code	Infrastructure as Code is the process of managing and provisioning computer data centers through machine-readable definition files, rather than physical hardware configuration or interactive configuration tools. Software examples include Terraform, Bicep, and CloudFormation.

Appendix B Cloud Offensive Toolkits

B.1 Azure

- MicroBurst - GitHub - NetSPI/MicroBurst: A collection of scripts for assessing Microsoft Azure security
- PowerZure - GitHub - hausec/PowerZure: PowerShell framework to assess Azure security

B.2 AWS

- weirdAAL - GitHub - carnal0wnage/weirdAAL: WeirdAAL (AWS Attack Library)
- pacu - GitHub - RhinoSecurityLabs/pacu: The AWS exploitation framework, designed for testing the security of Amazon Web Services environments.

B.3 GCP

- gcpHound - gcpHound : A Swiss Army Knife Offensive Toolkit for Google Cloud Platform (GCP)
- GCPBucketBrute - GitHub - RhinoSecurityLabs/GCPBucketBrute: A script to enumerate Google Storage buckets, determine what access you have to them, and determine if they can be privilege escalated.

---

[1] "SigmaHQ/sigma: Generic Signature Format for SIEM Systems," [Online]. Available: https://github.com/SigmaHQ/sigma. [Accessed 5 July 2022].
[2] SigmaHQ/pySigma: Python library to parse and convert Sigma rules into queries (and whatever else you could imagine),"[Online]. Available: https://github.com/SigmaHQ/pySigma#backends-comparison-between-pysigma-and-sigmac. [Accessed 5 July 2022].
[3] MITRE ATT&CK is a globally accessible knowledge base of adversary tactics and techniques based on real-world behavior. Available: https://attack.mitre.org/
[4] Sigma, [Online]. Available: https://github.com/SigmaHQ/sigma. [Accessed 31 March 2022].
[5] Specification: Sigma Correlations," https://github.com/SigmaHQ/sigma/wiki/Specification:-Sigma-Correlations. [Accessed 5 July 2022].
[6] "Compare AWS and Azure services to Google Cloud," [Online]. Available: https://cloud.google.com/free/docs/aws-azure-gcp-service-comparison. [Accessed 5 July 2022].
[7] Palacin and Safari, *Practical Threat Intelligence and Data-Driven Threat Hunting*.
[8] ATT&CK Cloud, [Online]. Available: https://attack.mitre.org/matrices/enterprise/cloud/. [Accessed 5 July 2022].

[9] Modify Cloud Compute Infrastructure, Technique T1578 - Enterprise | MITRE ATT&CK®, [Online]. Available: https://attack.mitre.org/techniques/T1578/. [Accessed 22 July 2022].
[10] "Configure anonymous public read access for containers and blobs," [Online]. Available: https://docs.microsoft.com/en-us/azure/storage/blobs/anonymous-read-access-configure?tabs=azure-cli [Accessed 5 July 2022].
[11] https://cloud.google.com/storage/docs/access-control/making-data-public
[12] Palacin and Safari, *Practical Threat Intelligence and Data-Driven Threat Hunting*.
[13] sigma - Visual Studio Marketplace
[14] "AWS Customer Support Policy for Penetration Testing," [Online]. Available: https://aws.amazon.com/security/penetration-testing/. [Accessed 31 March 2022].
[15] "Google Cloud Platform Acceptable Use Policy," [Online]. Available: https://cloud.google.com/terms/aup/. [Accessed 31 March 2022].
[16] "Google Cloud Platform Terms of Service," [Online]. Available: https://cloud.google.com/terms/. [Accessed 31 March 2022].
[17] Microsoft Cloud, "Penetration Testing Rules of Engagement," [Online]. Available: https://www.microsoft.com/en-us/msrc/pentest-rules-of-engagement?rtc=1. [Accessed 31 March 2022].